DOCUMENT RESUME

ED 309 186 TM 013 657

AUTHOR Kogut, Jan
TITLE Asymptotic Distribution of an IRT Person Fit Index. Project Psychometric Aspects of Item Banking No. 38. Research Report 88-13.
INSTITUTION Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE Nov 88
NOTE 41p.
AVAILABLE FROM Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Ability; Computer Simulation; Foreign Countries; *Item Banks; *Latent Trait Theory; Test Construction; *Test Items
IDENTIFIERS Aberrance; *Asymptotic Distributions; *Person Fit Measures; Rasch Model; Residuals (Statistics); Subtests

ABSTRACT

The distribution of a certain item response theory (IRT) based person fit index to identify systematic types of aberrance is discussed. For the Rasch model, it is proved that: (1) the joint distribution of subtest-residuals (the components of the index) is asymptotically multivariate normal; and (2) the distribution of the index is asymptotically chi-square. The parameters of these asymptotic distributions depend on whether ability of a person is known or estimated. Furthermore, the rate of convergence to the asymptotic distribution of the subtest-residuals is analyzed. In order to verify the results for short tests, a simulation study was conducted. The hypothetical test was composed of 40 items designed according to the Rasch model. Four data tables and two graphs present the numerical data from the simulation. (Author/SLD)

Asymptotic Distribution of an IRT Person Fit Index

Research Report 88-13

Jan Kogut

department of EDUCATION

Division of Educational Measurement and Data Analysis

University of Twente

Project Psychometric Aspects of Item Banking No.38

Colofon:
Typing: Lorette Padberg
Cover design: Audiovisuele Sectie TOLAB Toegepaste Onderwijskunde
Printed by: Centrale Reproductie-afdeling

Asymptotic Distribution of an IRT Person Fit Index

Jan Kogut

Asymptotic distribution of an IRT person fit index / Jan Kogut - Enschede : University of Twente, Department of Education, November, 1988. - 31 pages

Abstract

The distribution of a certain IRT based person fit index to identify systematic types of aberrance is discussed. For the Rasch model it is proved that: (1) the joint distribution of subtest-residuals (the components of the index) is asymptotically multivariate normal, and (2) the distribution of the index is asymptotically chi-square. The parameters of these asymptotic distributions depend on whether ability of a person is known or estimated. Furthermore, the rate of convergence to the asymptotic distribution of the subtest-residuals is analyzed. In order to verify the results for short tests, a simulation study is conducted.

Asymptotic Distribution of an IRT Person Fit Index

Introduction

If an item response theory (IRT) model correctly describes the performance of a population of persons on a test, the item parameters are known and certain regularity conditions are satisfied, then the maximum likelihood (ML) estimators of ability are consistent, asymptotically normally distributed, and efficient (Hambleton & Swaminathan, 1985, chapters 5 and 7; Lord, 1983). Yet, these properties are no longer obvious in the presence of occasional or systematical deviations of the person's response behavior from the model, which are likely to arise in educational measurement and testing (Fyans, 1982). When *occasional* measurement disturbances, such as temporary carelessness/guessing are present, robust estimation of ability may be a good solution. Robust ability estimates show—at the cost of only a slightly increased estimation error for the regular patterns—a significantly decreased estimation error for the aberrant patterns (Jones, 1982; Mislevy & Bock, 1982; Wainer & Wright, 1980). However, if *systematic* measurement disturbances, like unfamiliarity with the specific domain or copying/guessing to complete the test are expected, then it is more appropriate to detect and to diagnose those deviations rather than to correct them by robust estimation.

For the detection of systematic aberrance, several person fit indices have been proposed (Levine & Drasgow, 1983; Smith, 1985; Trabin & Weiss, 1983). Unfortunately, the exact or asymptotic null distribution of these indices are not known and therefore their use is limited.

In this paper, the asymptotic distribution of a simple modification of Smith's (1985) person fit index is investigated. For the case of known ability, the asymptotic distributions of subtest-residuals (the components of the index) as well as of the index itself are derived for a general IRT model. Next, still within the general framework of IRT, a basic system of equations connecting subtest-residuals both for known and ML estimated ability is found. Subsequently, the asymptotic distributions of the subtest-residuals and of the index when ability is estimated by the ML method are obtained, however, for the Rasch model only. Furthermore, for the subtest-residuals, the rate of convergence to the asymptotic distribution is evaluated.

Person Fit Index for Diagnosing Aberrance

Consider a test of n dichotomously scored items, and let $P_i \equiv P_i(\theta) = P(X_i=1|\theta)$ denote the probability of a correct response to item i $(i = 1,2,\ldots,n)$ for a person with ability θ. As usual in IRT, it will be assumed that the P_i's are increasing functions of θ, and that their first three derivatives with respect to θ exist and are finite for all values of θ. Several examples of such functions, including the well-known one-, two- and three- parameter logistic ones, can be found in Hambleton and Swaminathan (1985, chapter 3). A person's response pattern will be denoted by $\mathbf{X} \equiv (X_1, X_2, \ldots, X_L)$. The responses are regarded as independent random variables given ability level θ, i.e., the local independence is assumed. Finally, it is assumed that all item parameters are known.

The ML estimate of ability θ, $\hat{\theta}$, for a person with the response pattern $\mathbf{X}$, is a solution of the likelihood equation

$$(1) \qquad \Sigma_i \frac{(X_i - \hat{P}_i)\hat{P}_i'}{\hat{P}_i \hat{Q}_i} = 0$$

(Hambleton & Swaminathan, 1985, chapter 5; Lord, 1983), where P_i' is the derivative of P_i with respect to θ, $Q_i \equiv 1-P_i$, and a hat above a function indicates that this function has to be calculated at $\hat{\theta}$. Lord (1983) found that for the three-parameter logistic model with known item parameters the ML estimate of ability is:

(2) consistent, i.e., $\hat{\theta} \xrightarrow{P} \theta$ as $n \rightarrow \infty$

and

(3) asymptotically normally distributed and efficient, more precisely, $\dfrac{\hat{\theta} - \theta}{I^{-\frac{1}{2}}(\theta)} \xrightarrow{d} N(0,1)$ as $n \rightarrow \infty$;

where

$$(4) \qquad I(\theta) \equiv \Sigma_i P_i'^2 / P_i Q_i$$

is the test information function, and $\xrightarrow{p}$ and $\xrightarrow{d}$ denote the convergence in probability and in distribution, respectively. In order to carry out the proof Lord needed the following assumptions: (1) θ is bounded, (2) the item difficulty and discrimination parameters are bounded, (3) the pseudo-guessing parameters are bounded away from one, and (4) the test is lengthened by adding strictly parallel forms. If person's response variables are non-identically distributed, then Lord's assumptions satisfy the regularity conditions of Bradley and Gart (1962) by which efficient ML estimation of ability is possible. In this paper we will use similar but more general assumptions, namely that there exist: a fixed ability interval [a,b], fixed constants $\epsilon_1, \epsilon_2, \epsilon_3 > 0$, and a fixed constant $C > 0$ such that

(i) $\quad \epsilon_1 \leq P_i(\theta) \leq 1-\epsilon_2 \quad$ for all $\theta \in [a,b]$,
and all $i=1,2,\ldots,n$;

and

(ii) $\quad \epsilon_3 \leq P_i'(\theta) \leq C \quad$ for all $\theta \in [a,b]$
and all $i=1,2,\ldots,n$

Note that under assumptions (i) and (ii), for each $\theta \in [a,b]$,

$$(5) \qquad I(\theta) \to \infty \quad \text{as } n \to \infty.$$

Furthermore, if (3) holds, then (2) is true if and only if (5) is satisfied. Hence, (5) is a necessary condition for (2) and (3) to hold together. Also, by arguments similar to those of Lord (1983), it can be shown that the ML estimators of ability, under assumptions (i) and (ii), satisfy (2) and (3) for each $\theta \in [a,b]$.

In this paper, the following between-subtests person fit index is used:

$$(6) \qquad BF \equiv \Sigma_j \frac{(\Sigma_{i \in S_j} X_i - \mu_j)^2}{\sigma_j^2},$$

where

$$(7) \qquad \begin{aligned} \mu_j &\equiv E(\Sigma_{i \in S_j} X_i) = \Sigma_{i \in S_j} P_i, \\ \sigma_j^2 &\equiv Var(\Sigma_{i \in S_j} X_i) = \Sigma_{i \in S_j} P_i Q_i, \end{aligned}$$

and S_j $(j = 1,2,\ldots,J)$ indicates disjoint subtests obtained from a partition of the original test. The BF index is a multiplication of Smith's (1985) UB index by a constant: $BF = (J-1)UB$. Note that the BF value for a given pattern is the sum of the squared standardized residuals on the S_j subtests (in short, subtest-residuals). Extreme (positive) values of the index indicate patterns with large deviations from the IRT expectations on the specific subtests. To detect such aberrant patterns Smith (1985, 1986) uses a posteriori fixed critical value justified by a simulation study (Smith, 1988).

Extreme values of the subtest-residuals show that aberrant behavior occurs on these subtests. Knowing the type of items in the subtests, attempts to diagnose the kind of aberrance may be made. Many person fit analyses are possible dependent on the manner of grouping items into subtests. For instance, items could be grouped into subtests according to: the increasing difficulty parameter of the items, their position in the test, the type of items, or the conceptual domains covered by the items. Each grouping may have a different diagnostic meaning. For example, a high positive value of the subtest-residual on the most difficult items may suggest that the person was guessing. Likewise, a high negative value of the subtest-residual on the item subset covering a specific domain may be an indication that the person has a poor knowledge of the domain. However, many person fit analyses should be conducted before a specific type of aberrance can be pointed out.

Asymptotic Distribution of the Index

Let us consider the asymptotic distribution of the subtest-residuals as well as of the index if the number of items tends to infinity. We will investigate two separate cases: (1) the person's ability θ is known, and (2) θ is estimated by the ML method from his/her response pattern

First, let us consider the case where ability is known. In this case, the subtest-residuals will be denoted by

$$(8) \qquad Y_j \equiv \frac{\Sigma_{i\in Sj} X_i - \mu_j}{\sigma_j}, \qquad \mathbf{Y} \equiv (Y_1, Y_2, \ldots, Y_J).$$

Since Y_j is the standardized sum of independent random variables $\{X_i,\ i\in S_j\}$ (by the assumption of local independence), we may apply the central limit theorem with the Liapunov condition (e.g. Rao, 1965, p.107). So:

$$(9) \qquad \begin{aligned} &\mu_i \equiv E(X_i) = P_i, \qquad \sigma_i^2 \equiv E(|X_i-\mu_i|^2) = P_iQ_i, \\ &\beta_i \equiv E(|X_i-\mu_i|^3) = P_iQ_i(P_i^2+Q_i^2) \le P_iQ_i; \end{aligned}$$

and

$$L_j \equiv \frac{(\Sigma_{i\in Sj}\beta_i)^{1/3}}{(\Sigma_{i\in Sj}\sigma_i^2)^{1/2}} \le (\Sigma_{i\in Sj}P_iQ_i)^{-1/6}.$$

If the number of items in S_j, n_j, tends to infinity, then under assumption (i), the series $\Sigma_{i\in Sj}P_iQ_i$ is divergent and, consequently, the Liapunov condition $\lim_{nj\to\infty} Lj = 0$ is satisfied and Yj has asymptotically the standard normal distribution. Furthermore, the random variables $Y_1, Y_2, \ldots, Y_J$, being functions of disjoint subsets of the independent random variables X_i's, are independent, too. Hence, we can conclude that the random vector $(Y_1, Y_2, \ldots, Y_J)$ has asymptotically the non-singular standard normal distribution of rank J, i.e.,

$$(10) \qquad (Y_1, Y_2, \ldots, Y_J) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}) \quad \text{as } n_j \to \infty \ (j=1,2,\ldots,J),$$

where $\mathbf{0}$ is a vector of J zero's (the vector of the Y_j's means) and $\mathbf{I}$ is a JxJ identity matrix (the covariance matrix of the Y_j's).

Now, let us investigate the asymptotic distribution of the BF index. Note that BF can be represented as a quadratic form in the variables $Y_1, Y_2, \ldots, Y_J$, namely, $BF = \Sigma_j Y_j^2 = \mathbf{YIY}'$, where $\mathbf{Y}'$ is the transpose of $\mathbf{Y}$. So, by (10) it becomes clear that $BF \xrightarrow{d} \mathbf{ZIZ}'$, where $\mathbf{Z}$ is $N(\mathbf{0}, \mathbf{I})$ distributed (Serfling, 1980, p.25, Corollary). Subsequently, we can conclude that BF is asymptotically chi-squared with J degrees of freedom, i.e.,

$$(11) \qquad BF \xrightarrow{d} \chi^2_J \quad \text{as } n_j \to \infty \quad (j=1,2,\ldots,J)$$

(Serfling, 1980, p.128, Lemma).

Second, let us consider the asymptotic distribution of the subtest-residuals and of the index when the exact value of θ is replaced by its ML estimate $\hat{\theta}$. In this case we will denote

(12) $\hat{Y}_j \equiv \frac{\Sigma_{i\epsilon S_j} X_i - \hat{\mu}_j}{\hat{\sigma}_j}, \qquad \hat{\mathbf{Y}} \equiv (\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_J)$

and

(13) $\hat{BF} \equiv \Sigma_j \hat{Y}_j^2 = \hat{\mathbf{Y}}\mathbf{I}\hat{\mathbf{Y}}'$,

where $\hat{\mu}_j$ and $\hat{\sigma}_j$ are calculated according to (7) at $\hat{\theta}$. Note that if the ML ability estimates are used instead of true values, then due to (1), $\hat{\theta}$, and hence, the $\hat{\mu}_j$'s and $\hat{\sigma}_j$'s are functions of the X_i's. Therefore, $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_J$ are dependent random variables. The consequence of the dependency can easily be found for the Rasch model (RM). For this model, where $P_i = 1/(1+\exp[-A(\theta-b_i)])$, ($b_i$ is the difficulty of item i and A is a positive constant), we have $P_i' = P_iQ_i$, and from (4) for the subtest information function we obtain

(14) $I_j(\theta) \equiv \Sigma_{i\epsilon S_j} P_i'^2/P_iQ_i = \Sigma_{i\epsilon S_j} P_iQ_i = \sigma_j^2$.

Putting this in (1) we thus have

(15) $\Sigma_i (X_i - \hat{P}_i) = \Sigma_j \hat{Y}_j \hat{I}_j^{1/2} = 0$.

Hence, for the RM, the random variables $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_J$ must satisfy *constraint* (15). For J = 2, this means that $\hat{Y}_1$ and $\hat{Y}_2$ are perfectly correlated. The subtest-residual over all items (if $J = 1$) is always equal to zero. In this way, similar results for other IRT models can be derived. Yet, as the dependency of $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_J$ results in a *constraint* we may expect that the joint distribution of the $\hat{Y}_j$'s is asymptotically a singular normal distribution of rank $J-1$, and accordingly $\hat{BF}$ is asymptotically chi-square distributed with $J-1$ degrees of freedom. Let us verify these hypotheses.

In order to find the joint asymptotic distribution of $(\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_J)$, the following system of equations derived from the definitions of $\hat{Y}_j$ and Y_j can be used

$$(16) \qquad \hat{Y}_j = Y_j - \frac{\hat{\mu}_j - \mu_j}{\sigma_j} + (\Sigma_{i \in S_j} X_i - \hat{\mu}_j)(\frac{1}{\hat{\sigma}_j} - \frac{1}{\sigma_j}),$$

where $j = 1, 2, \ldots, J$. Let us start with the question how the term $(\hat{\mu}_j - \mu_j)/\sigma_j$ can be approximated. Note that

$$(17) \qquad \sigma_j{}^2 = n_j(\bar{P}_j\bar{Q}_j - s_j{}^2),$$

where

$$(18) \qquad \bar{P}_j \equiv (1/n_j)\Sigma_{i \in S_j} P_i, \quad s_j{}^2 \equiv (1/n_j)\Sigma_{i \in S_j}(P_i - \bar{P}_j)^2,$$

and $\bar{Q}_j \equiv 1 - \bar{P}_j$. Since under assumptions (i) and (ii), (3) holds and from (4) $I^{-½}(\theta) = O(n^{-½})$, we have

(19) $\quad \hat{\theta} - \theta = Op(n^{-1/2}),$

where $U_n = Op(v_n)$ denotes that for every $\epsilon>0$ there exists $K_\epsilon>0$ and N_ϵ such that $P(|U_n/v_n| \leq K_\epsilon) \geq 1-\epsilon$ for all $n>N_\epsilon$, i.e., U_n/v_n is bounded in probablity. Now, applying the Taylor expansion to $\hat{P}_i$, and using (19) we obtain

$$\hat{P}_i - P_i = (\hat{\theta} - \theta)P_i' + op(n^{-1/2}),$$

where $U_n = op(v_n)$ denotes $U_n/v_n \xrightarrow{P} 0$. Summing both sides of the equation over $i \epsilon S_j$, dividing them by σ_j, and applying (17), we obtain

$$(20) \qquad \frac{\hat{\mu}_j - \mu_j}{\sigma_j} = \frac{(\hat{\theta} - \theta)}{\sigma_j} \Sigma_{i\epsilon Sj} P_i' + op(1).$$

Likewise, it can be shown that

$$\frac{1}{\sigma_j}\Sigma_i \frac{(X_i-\hat{P}_i)\hat{P}_i'}{\hat{P}_i\hat{Q}_i} = \frac{1}{\sigma_j}\Sigma_i \frac{(X_i-P_i)P_i'}{P_iQ_i} + \frac{(\hat{\theta}-\theta)}{\sigma_j}\frac{d}{d\theta}\{\Sigma_i \frac{(X_i-P_i)P_i'}{P_iQ_i}\}$$

(21)

$$+ op(1).$$

Having calculated the derivative $d\{\Sigma_i(X_i-P_i)P_i'/P_iQ_i\}/d\theta$, it is then readily seen that the derivative may be written as $[-I(\theta)+\Sigma_i K_i(X_i-P_i)]$, where the K_i's do not depend on X_i and are bounded (again under assumptions (i) and (ii) and because all P_i'' were assumed to be finite). Calculating the mean and variance of the last sum and using the Chebyshev inequality, we obtain $\Sigma_i K_i(X_i - P_i) = O_P(n^{1/2})$. It is then easy to show, by the last result, (19) and (17) that $[(\hat{\theta}-\theta)/\sigma_j][\Sigma_i K_i(X_i-P_i)] = o_P(1)$. Using this relation and (1) in (21), we have

$$(22) \qquad \frac{\hat{\theta} - \theta}{\sigma_j} = \frac{1}{I(\theta)\sigma_j} \Sigma_i \frac{(X_i - P_i)P_i'}{P_i Q_i} + o_P(1).$$

Finally, applying the Chebyshev inequality, (20) and (17) in the similar way, it can be shown that

$$(23) \qquad (\Sigma_{i \in S_j} X_i - \hat{\mu}_j)(\frac{1}{\hat{\sigma}_j} - \frac{1}{\sigma_j}) = o_P(1).$$

Using the results (20), (22) and (23), the system of equations (16) can be replaced by the following one

$$(24) \qquad \hat{Y}_j = Y_j - \frac{\Sigma_{i \in S_j} P_i'}{I(\theta)\sigma_j} \Sigma_i \frac{(X_i - P_i)P_i'}{P_i Q_i} + o_P(1), \quad (j=1,\ldots,J)$$

which will be the basis of our further considerations.

Now, let us return to the RM where, $P_i' = P_i Q_i$ and (14) holds. For this model, the system of equations (24) may be expressed in a simple matrix notation. Denoting

$$\mathbf{R} \equiv ([I_1(\theta)/I(\theta)]^{1/2}, [I_2(\theta)/I(\theta)]^{1/2}, \ldots, [I_J(\theta)/I(\theta)]^{1/2}),$$

and using (14) and an analog of the first equality in (15) at θ, we obtain from (24) the following important matrix equation

$$(25) \qquad \hat{\mathbf{Y}}' = \mathbf{Y}' - \mathbf{R}'\mathbf{R}\mathbf{Y}' + \mathbf{z}' = \mathbf{A}\mathbf{Y}' + \mathbf{z}',$$

where $\mathbf{A} \equiv \mathbf{I}-\mathbf{R}'\mathbf{R}$, and $\mathbf{z} \equiv (z_1, z_2 \ldots, z_J)$ is a vector of $o_p(1)$ random variables which converges in probability to zero. Let us now assume that

$$(\text{iii}) \qquad \frac{I_j(\theta)}{I(\theta)} \rightarrow r_j \quad \text{as } n_j \rightarrow \infty \ (j=1,2,\ldots,J).$$

Then

$$(26) \qquad \mathbf{R} \rightarrow \mathbf{R}_o \equiv (r_1^{1/2}, r_2^{1/2}, \ldots, r_J^{1/2}), \qquad \mathbf{A} \rightarrow \mathbf{A}_o \equiv \mathbf{I}-\mathbf{R}_o'\mathbf{R}_o,$$

and because of the identity $\Sigma_j I_j(\theta) = I(\theta)$, $\Sigma_j r_j = 1$ and

$$(27) \qquad \mathbf{R}_o\mathbf{R}_o' = 1$$

So, from the multivariate version of Slutsky's theorem (Rao, 1965, p.102, Corollary (x)), we can conclude from (25) and (26) that

(28) $\hat{\mathbf{Y}}' \xrightarrow{d} \mathbf{A}_0\mathbf{Y}'$.

Subsequentely, by (10) we have $\hat{\mathbf{Y}}' \xrightarrow{d} N(0,\mathbf{A}_0\mathbf{I}\mathbf{A}_0')$, (Serfling, 1980, p. 26, Application A). But $\mathbf{A}_0\mathbf{I}\mathbf{A}_0' = \mathbf{I}-\mathbf{R}_0'\mathbf{R}_0$, because $\mathbf{A}_0 = \mathbf{A}_0'$ and (27) holds. This means also that the matrix $\mathbf{I}-\mathbf{R}_0'\mathbf{R}_0$ is idempotent, and thus of rank $r=\text{rank}\,(\mathbf{I}-\mathbf{R}_0'\mathbf{R}_0) = \text{trace}\,(\mathbf{I}-\mathbf{R}_0'\mathbf{R}_0) = \Sigma_j(1-r_j) = J-1$. In turn as known, the characteristic function of a $N(0,\mathbf{I}-\mathbf{R}_0'\mathbf{R}_0)$ distributed variable $\mathbf{Z}$ is $\phi(t) = e^{-\frac{1}{2}t(\mathbf{I}-\mathbf{R}_0'\mathbf{R}_0)t'}$. Hence, by an orthogonal transformation $\mathbf{T}'=\mathbf{C}t'$ such that $T_j = \Sigma_j t_j r_j^{\frac{1}{2}}$, we can obtain $Q(t) = Q(\mathbf{T}) = \Sigma_{j=1}^{J-1} T_j^2$ (Rao, 1965, chapter 3b3). This means that the total mass of $\mathbf{Z}$ is situated in the hyperplane $\Sigma_j z_j r_j^{\frac{1}{2}}=0$. So we may regard the following theorem as proved:

<u>Theorem</u>. If for the Rasch model assumptions (i), (ii) and (iii) are satisfied and ability is estimated by the maximum likelihood method, then the joint distribution of the subtest-residuals is asymptotically a singular normal distribution of rank J-1, i.e.,

(29) $(\hat{Y}_1,\hat{Y}_2,\ldots,\hat{Y}_J) \xrightarrow{d} N(0,\mathbf{I}-\mathbf{R}_0'\mathbf{R}_0)$

as $n_j \to \infty$ $(j=1,2,\ldots,J)$,

the total mass of which is situated in the hyperplane $\Sigma_j z_j r_j^{\frac{1}{2}} = 0$.

From the theorem it immediately follows that the asymptotic variance and correlation of the $\hat{Y}_j$ subtest-residuals are

$$(30) \qquad \mathrm{Var}(\hat{Y}_j) \to 1-r_j, \quad \mathrm{Corr}(\hat{Y}_j, \hat{Y}_k) \to \frac{-(r_j r_k)^{1/2}}{\{(1-r_j)(1-r_k)\}^{1/2}}$$

as n_j, $n_k \to \infty$. If $J = 2$, then $\mathrm{Corr}(\hat{Y}_j, \hat{Y}_k) \to -1$. This corresponds to the result on page 10, where the correlation has been shown to be perfect. Considering also that $\Sigma_j z_j r_j^{1/2}=0$ corresponds to (15), the asymptotic distribution of the $\hat{Y}_j$'s possess the same two properties as their exact distribution.

Having the limiting distribution of the subtest-residuals, the limiting distribution of $\hat{BF}$ can be considered. By (13) and (28) we have $\hat{BF} \xrightarrow{d} \mathbf{Y}\mathbf{A}_O'\mathbf{I}\mathbf{A}_O\mathbf{Y}'$ (Serfling, 1980, p.25, Corollary), where $\mathbf{A}_O'\mathbf{I}\mathbf{A}_O = \mathbf{I}-\mathbf{R}_O'\mathbf{R}_O$ is, as was mentioned, idempotent and of rank J-1. Applying the Fisher-Cochran theorem (Serfling, 1980, p.128, Lemma), we may conclude:

<u>Corollary</u>. If for the Rasch model assumptions (i), (ii) and (iii) are satisfied and ability is estimated by the maximum likelihood method, then the index is asymptotically chi-square distributed with J-1 degrees of freedom, i.e.,

$$(31) \qquad \hat{BF} \xrightarrow{d} \chi^2_{J-1} \quad \text{as } n_j \to \infty \ (j=1,2,\ldots,J).$$

Results (10), (11), (29) and (31) may be very useful in applications, provided that the error of the approximations for tests of realistic length is not too large.

Let us restrict ourselves to the error of the standard normal approximation for the subtest-residuals when ability is known (see (10)). For any independent random variables X_i's that have finite third absolute moments, the following inequality holds

$$(32)\qquad D_j \equiv \sup_x |F_j(x)-\Phi(x)| \le C\,\frac{\Sigma_{i\in S_j}E(|X_i-\mu_i|^3)}{\{\Sigma_{i\in S_j}E(|X_i-\mu_i|^2)\}^{3/2}} \equiv B_j,$$

where $F_j(x)$ is the distribution of the n_j standardized summands, $\Phi(x)$ is the standard normal distribution, and C is an universal constant independent of n_j and of any characteristic of the X_i's (Serfling, 1980, p.33). It is easy to show that if our specific X_i's are identically distributed ($P_i = P$ and thus $Q_i = Q$ for all $i \in S_j$), then the order of D_j, which is given by the inequality, *cannot* be improved. From the classical theorem of de Moivre-Laplace it follows that the function $F_j(x)$ (here the distribution of Y_j) is discontinuous at the points $x_k=(k-n_jP)/(n_jPQ)^{½}$, ($k=0,1,\ldots,n_j$), with jumps asymptotically equal to $\{1/(2\pi n_jPQ)^{½}\}\exp(-x_k^2/2)$. Hence D_j is of order $O(n_j^{-½})$. On the other hand, using (9), we may conclude that B_j is of the same order. Thus, in general, for the evaluation of the error it is sufficient to consider the B_j bound in (32).

Substituting in (32) the moments of X_i, being calculated in (9), applying $P_i^2+Q_i^2 = 1-2P_iQ_i$ and the Cauchy-Schwarz inequality $(1/n_j)(\Sigma_{i\epsilon S_j}P_iQ_i)^2 \le \Sigma_{i\epsilon S_j}(P_iQ_i)^2$, and then using (17) we obtain

$$(33) \qquad D_j \le B_j \le \frac{C}{\{n_j(\bar{P}_j\bar{Q}_j-s_j^2)\}^{½}}[1-2(\bar{P}_j\bar{Q}_j-s_j^2)].$$

Again with $P_i=P$ for all $i\epsilon S_j$, it is easy to show that the second inequality *cannot* be improved. So, the bound for D_j error of the standard normal approximation for Y_j in (10), is a function of n_j as well as of the variability of the subtest $P_i(\theta)$'s at a fixed θ, $\bar{P}_j$ and s_j^2. The inequalities show that the bound for D_j, B_j, mainly depends on the behavior of $C/\{n_j(\bar{P}_j\bar{Q}_j-s_j^2)\}^{½}$. Therefore, for a fixed $\bar{P}_j$ and s_j^2, B_j is of order $O(n_j^{-½})$. Next, for a fixed n_j, the following conclusions can be drawn: (a) B_j is minimal when at θ all $P_i(\theta) = 1/2$ (note that in this case $D_j \le C/(0.25n_j)^{½}$); (b) if the average of the $P_i(\theta)$'s at θ, $\bar{P}_j$, tends to 0 or 1, then B_j tends to infinity; (c) at a fixed average $\bar{P}_j$, the larger the variation of the $P_i(\theta)$'s, s_j^2, the larger B_j. These conclusions hold for an arbitrary IRT model for which (10) is satisfied.

For the RM in particular, using (14) and (17) we obtain

$$(34) \qquad I_j(\theta) = n_j(\bar{P}_j\bar{Q}_j-s_j^2),$$

and from (33)

$$(35) \qquad D_j \le B_j \le \frac{C}{I_j^{½}} (1 - 2I_j/n_j).$$

The behavior of the subtest information I_j is illustrated in Figure 1(a) for two cases of S_j subtest: (1) of items with similar difficulties, and (2) of items with distant difficulties. However, in both cases the means of the difficulties are equal, i.e., $\bar{b}_{j(1)} = \bar{b}_{j(2)} = b_o$. Then for ability θ very close to b_o (where $P_j \approx ½$), $I_{j(1)} > I_{j(2)}$ because, as can be concluded from the $P_i(\theta)$'s for the RM, $P_{j(1)} \approx P_{j(2)}$ while $s_j^2{}_{(1)} < s_j^2{}_{(2)}$. Yet, for more extreme θ's, $I_{j(1)} < I_{j(2)}$ because $P_{j(1)} < P_{j(2)}$ while $s_j^2{}_{(1)} \approx s_j^2{}_{(2)}$. Of course, for $\theta \to \pm \infty$, $Ij(1)$ and $Ij(2)$ tend to zero irrespective of the difficulties of the items. If $P_i(\theta) = P(\theta)$ for all $i \in S_j$, then $b_i = b$ and the highest possible information is obtained at $\theta = b$ which is equal to $0.25n_j$. According to the behavior of I_j and (35), the bound for D_j error in the two cases is illustrated in Figure 1(b). The minimal possible error is obtained at $\theta = b$ when $P_i(\theta) = P(\theta)$ for all $i \in S_j$. However, for the subtest of items with more spreaded difficulties, the standard normal approximation for the subtest-residual is better in a broader range of ability. With increasing the number of items in the subtest, the approximation becomes better and it is of order $n^{-½}$.

Numerical Results

In order to examine the degree of approximation by the asymptotic distributions, a hypothetical test was designed according to the RM. The test was composed of 40 items with the item difficulties b_i sampled from the normal distribution with mean 0.00 and variance 1.56^2. To investigate the effect of different grouping of items into subtests, all items were ordered and numbered corresponding to the values of b_i (from the lowest to the highest) and then divided into a few subtests in three analyses. In the *first* analysis, two subtests of 20 items were formed (the first subtest consisted items with numbers from 1 to 20; the second subtest consisted items 21 to 40). In the *second* analysis, four subtests of 10 items (items 1 to 10, items 11 to 20, etc.) and in the *third* analysis eight subtests of 5 items (items 1 to 5, items 6 to 10, etc.) were constructed. The parameters $\bar{P}_j$, s_j^2 and I_j of these subtests, calculated using (18) and (34) for ability $\theta = 0.00$, are presented in Table 1.

Insert Table 1 about here

Subsequently, five hundred patterns were generated for $\theta = 0.00$ according to the RM and the values of the subtest-residuals and of the index were calculated in the three analyses, as well.

The mean and variance of the empirical distribution of the subtest-residuals are given in columns 2 and 3 of Tables 2, 3 and 4, and the mean and variance of the empirical

Insert Tables 2, 3 & 4 about here

distribution of the index in columns 6 and 7, both for the case of known ability (i.e., $\theta = 0.00$) and the case of ability estimated by the ML method (applying the PML program, see Gustafsson, 1981). The corresponding results for the asymptotic distributions, calculated using (10) and (11) for *known ability*, and using (29), (30) and (31) for the *ML estimated ability*, are given within brackets immediately below the empirical ones. As is easily seen, a significant agreement between the empirical and asymptotic means and variances was obtained.

Furthermore, in order to investigate whether the empirical distributions can be approximated by their asymptotic ones, the Kolmogorow-Smirnow and Chi-Square tests of goodness-of-fit were applied (with the help of the NPAR TESTS of the SPSS, see Hull and Nie, 1981).

The results of the Kolmogorow-Smirnow test for the distribution of subtest-residuals are given in columns 4 and 5 of Tables 2, 3 and 4. Here the theoretical distribution of the Kolmogorow-Smirnow test was assumed to be normal with mean and variance given within the brackets (i.e., the appropriate asymptotic distribution). In columns 4 there are listed values of D_j defined in (32). For the case of *known ability*, values of D_j are found to be proportional to the reciprocal of the root of I_j (see (35) and Table 1). In addition, the results confirm the rules (a), (b) and (c), specified on page 18. However, it seems that the same conclusions hold for the case of the *ML estimated ability*. Columns 5 of Tables 2, 3 and 4 show that the probability of exceedance for the Kolmogorow–Smirnow statistic (and thus also for the D_j statistic) will be higher than the observed one. The results obtained for known ability indicate that the distribution of the subtest–residuals is not satisfactory approximated by the standard normal distribution, at least in our example of 20, 10 and 5 item subtests. However, the results obtained for ML estimated ability seem to be more promising. For instance, in Table 3 two of the four distributions of $\hat{Y}_j$ can be regarded as normal at the 0.05 significance level. This discrepancy may be due to the fact that the number of possible values of the $\hat{Y}_j$ statistic is higher than that one of the Y_j statistic. Therefore, for estimated ability the empirical distribution is smoother and with lower values of D_j.

Finally, the results of the usual Chi–Square goodness–of–fit test for the distribution of the index are shown in the last three columns of Tables 2, 3 and 4. Here the data were pooled according to the 30%–, 50%–, 70%, 90%– and 95%–th percentiles of the appropriate chi–squared distribution (with the number of degrees of freedom according to (11) for the case of known ability, and according to (31) for ML estimated ability). Using these percentiles, empty classes could be avoided. As can be seen, the null hypoth‹ -i: of a chi-square distribution of the index cannot be rejected at the 0.05 significance level. In summary, the asymptotic distributions may certainly be useful for the purpose of person fit analyses, even for the tests of relatively low length.

Discussion

IRT offers several person fit indices to identify systematic types of aberrance in a person's response behavior. Unfortunately, the proper use of these indices has so far been limited by an insufficient knowledge of their exact or asymptotic null distribution. Therefore, the asymptotic distribution of the person fit index and of the subtest-residuals given in ıis paper should have some practical value, at least for the Rasch model and when ability is estimated by the ML method. So, for instance, in order to detect the type of aberrance called un-, or super-familiarity with specific domains, the following procedure can be applied. *First*, the subsets of items covering the J domains concerned should be specified. *Second*, for a given pattern, the ML ability estimate $\hat{\theta}$ and the values of $\hat{Y}_j$ and $\hat{BF}$ must be calculated. If the value of $\hat{BF}$ is higher than the 100*α% percentile of the χ^2 distribution with J-1 degrees of freedom (see Corollary), then the pattern is classified as aberrant with the Type I error equal to α . *Third*, in order to examine whether the detected aberrance is due to a particular domain, values of $\hat{Y}_j$ divided by its standard deviations $(1-\hat{I}_j/\hat{I})^{1/2}$, for each of J subsets (see (30) and (iii)), have to be calculated. If one or more of these values are lower than the 100*$(\alpha/2)$% percentilc or higher than the 100*$(1-\alpha/2)$% percentile of the standard normal distribution (see Theorem), then we may conclude an interaction between the subtest scores and the domains. So, we may suspect that a person is un-familiar with one domain, while he/she is super-familiar with another. Of course, many other types of aberrance should be tested before we definitively conclude to the type of aberrance involved in the test-taking behavior of the given person.

In order to determine the asymptotic distributions, assumptions (i), (ii) and (iii) have been made. These assumptions are rather technical than restrictive, thus, they can be easy realized when applying the Rasch model. Considering the fact that tests are usually designed for a given population of persons, they seem to be mild enough to satisfy all standard test settings. The *first assumption* stipulates that there must exist an interval of ability for which there are no items with the item characteristic curves (ICC's) closely approaching the zero and one asymptotes. As those items are just the ones which are too easy or too difficult to be included in the test for a given population, this assumption should easily be satisfied in practice. The *second assumption* requires that for the same ability interval there are no items of which ICC's are neither almost horizontal nor almost vertical. This assumption is also met in most standardized test settings as items of which ICC's are very flat are considered to be too little discriminating to be adopted in the test. Likewise, items of which the ICC's are nearly vertical simply do not exist in practice. The *last assumption* stipulates that, with increasing the number of items in the test, the ratio's of subtest and test information approach certain constants. In practice, this is easy to realize as usually in the construction of a test items are sampled from subdomains in accordance with a fixed distribution.

References

Bradley, R.A. & Gart J.J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. Biometrika, 49, 205-214.

Fyans, J.Jr. (Ed.) (1982). Achievement motivation: Recent trends in theory and research. New York: Plenum Press.

Gustafsson, J.E. (1981). PML: a computer program for conditional estimation and testing in the Rasch model for dichotomous items. Reports from the Institute of Education, University of Goteborg, no. 85.

Hambleton, R.K. & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston: Kluwer-Nijhoff Publishing.

Jones, D.H. (1982). Tools of robustness for item response theory. In D.J. Weiss (Ed.), Proceedings of the 1982 IRT and computerized adaptive testing conference. Minneapolis: University of Minnesota.

Levine, M.V., & Drasgow, F. (1983). Appropriateness measurement: validating studies and variable ability models. In D.J. Weiss (Ed.), New horizons in testing: latent trait theory and computerized adaptive testing. New York: Academic Press.

Lord, F.M. (1983). Unbiased estimates of ability parameters, of their variance, and of their parallel-forms reliability. Psychometrika, 48, 233-245.

Mislevy, R.J., & Bock, R.D. (1982). Biweight estimates of latent ability. Educational and Psycholcgical Measurement, 42, 725-737.

Hull, C.H., & Nie, N.H. (1981). SPSS update 7-9: new procedures and facilities for releases 7-9. New York: McGraw-Hill.

Rao, C.R. (1965). Linear statistical inference and its applications. New York: Wiley.

Serfling, R.J. (1980). Approximation theorems of mathematical statistics. New York: Wiley.

Smith, R.M. (1985). A comparison of Rasch person analysis and robust estimators. Educational and Psychological Measurement, 45, 433-444.

Smith, R.M. (1986). Person fit in the Rasch model. Educational and Psychological Measurement, 46, 359-372.

Smith, R.M. (1988). The distributional properties of Rasch standardized residuals. Educational and Psychological Measurement, 48, 657-667.

Trabin, E.T., & Weiss, D.J. (1983). The person response curve: fit of individuals to item response theory models. In D.J. Weiss (Ed.), New horizons in testing: latent trait theory and computerized adaptive testing. New York: Academic Press.

Wainer, H., & Wright, B.D. (1980). Robust estimation of ability in the Rasch model. Psychometrika, 45, 373-391.

Acknowledgements

The author is indebted to Ron J.H. Engelen, Wim J. van der Linden and Sebie J. Oosterloo for their comments on an earlier version of the paper. The content of the paper is however fully his own responsibility.

Table 1

Parameters of subtests for known ability ($\theta = 0.00$)

Subtest	Number of items	Subtest Parameter $\bar{P}_j$	s_j^2 $(*10^2)$	I_j
		First Analysis		
1	20	0.701	2.109	3.768
2	20	0.189	1.211	2.823
		Second Analysis		
1	10	0.825	0.467	1.397
2	10	0.576	0.669	2.371
3	10	0.282	0.556	1.968
4	10	0.096	0.143	0.855
		Third Analysis		
1	5	0.882	0.201	0.510
2	5	0.768	0.083	0.887
3	5	0.646	0.385	1.124
4	5	0.509	0.049	1.247
5	5	0.340	0.373	1.104
6	5	0.223	0.050	0.864
7	5	0.126	0.089	0.546
8	5	0.066	0.020	0.309

Table 2

Mean and variance of empirical and asymptotic distributions, and results of goodness-of-fit tests from first analysis

Sub test	Subtest-Residual		K-S Test		Index		Chi-Square Test		
	Mean	Var.	D_o	$P(Dn^{½} \geq D_o)$	Mean	Var.	χ^2_o	df	$P(\chi^2 \geq \chi^2_o)$
				Known Ability ($\theta = 0.00$)					
1	−0.03	0.96	0.127	0.000					
	(0.00)	(1.00)			1.90	3.60	7.617	2	0.179
2	0.03	0.94	0.136	0.000	(2.00)	(4.00)			
	(0.00)	(1.00)							
				ML Estimated Ability ($\theta = \hat{\theta}$)					
1	−0.02	0.42	0.073	0.010					
	(0.00)	(0.43)			0.97	1.86	8.727	1	0.120
2	0.02	0.55	0.061	0.050	(1.00)	(2.00)			
	(0.00)	(0.57)							

Note. The parameters of the asymptotic distributions are given within brackets.

Table 3

Mean and variance of empirical and asymptotic distributions, and results of goodness of fit tests from second analysis

Sub test	Subtest–Residual Mean	Subtest–Residual Var.	K–S Test D_o	K–S Test $P(Dn^{½} \geq D_o)$	Index Mean	Index Var.	Chi–Square Test χ^2_o	Chi–Square Test df	Chi–Square Test $P(\chi^2 \geq \chi^2_o)$
	Known Ability (θ = 0.00)								
1	0.03 (0.00)	0.97 (1.00)	0.177	0.000					
2	–0.06 (0.00)	1.00 (1.00)	0.144	0.000	3.87 (4.00)	7.32 (8.00)	5.637	4	0.343
3	0.02 (0.00)	0.94 (1.00)	0.160	0.000					
4	0.02 (0.00)	0.97 (1.00)	0.243	0.000					
	ML Estimated Ability ($\theta = \hat{\theta}$)								
1	0.07 (0.01)	0.81 (0.79)	0.080	0.003					
2	–0.03 (0.00)	0.62 (0.64)	0.058	0.070	2.95 (3.00)	5.65 (6.00)	10.187	3	0.070
3	–0.01 (0.00)	0.69 (0.70)	0.046	0.248					
4	–0.03 (0.00)	0.83 (0.87)	0.119	0.000					

Note. The parameters of the asymptotic distributions are given within brackets.

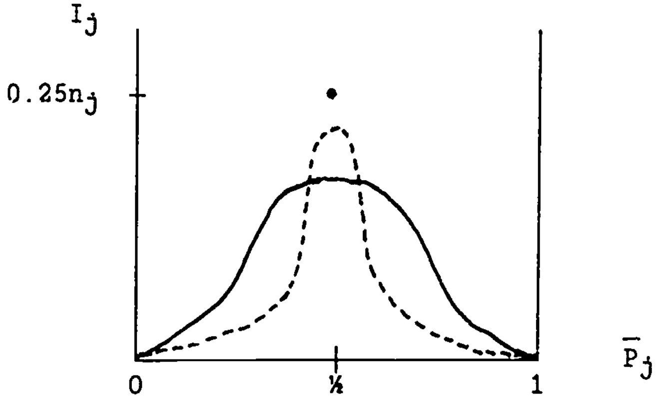
Table 4

Mean and variance of empirical and asymptotic distributions, and results of goodness of fit tests from third analysis

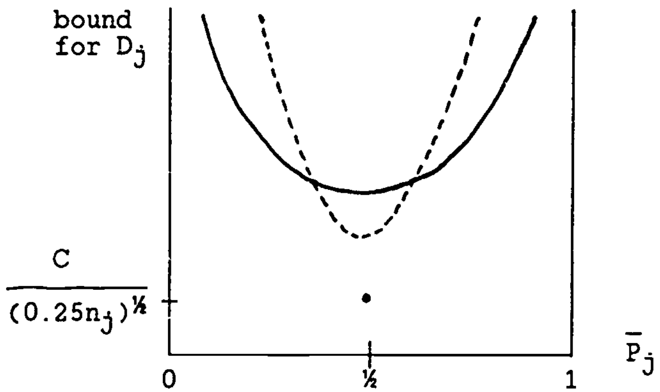
Sub test	Subtest-Residual Mean	Subtest-Residual Var.	K-S Test D_o	K-S Test $P(Dn^{½} \geq D_o)$	Index Mean	Index Var.	Chi-Square Test χ^2_o	df	$P(\chi^2 \geq \chi^2_o)$
			Known Ability (θ = 0.00)						
1	0.05 (0.00)	0.99 (1.00)	0.358	0.000					
2	−0.01 (0.00)	0.98 (1.00)	0.238	0.000					
3	−0.03 (0.00)	1.02 (1.00)	0.180	0.000					
4	−0.05 (0.00)	1.05 (1.00)	0.197	0.000	8.05 (8.00)	18.85 (16.00)	8.737	8	0.120
5	0.07 (0.00)	1.00 (1.00)	0.193	0.000					
6	−0.04 (0.00)	1.02 (1.00)	0.239	0.000					
7	−0.01 (0.00)	0.96 (1.00)	0.311	0.000					
8	0.05 (0.00)	1.02 (1.00)	0.403	0.000					
			ML Estimated Ability ($\theta = \hat{\theta}$)						
1	0.09 (0.00)	0.89 (0.92)	0.277	0.000					
2	0.03 (0.00)	0.89 (0.87)	0.083	0.002					
3	−0.00 (0.00)	0.86 (0.83)	0.059	0.062					
4	−0.04 (0.00)	0.87 (0.81)	0.066	0.027	7.13 (7.00)	16.65 (14.00)	2.727	7	0.742
5	0.05 (0.00)	0.90 (0.83)	0.063	0.039					
6	−0.07 (0.00)	0.89 (0.87)	0.135	0.000					
7	−0.04 (0.00)	0.92 (0.92)	0.230	0.000					
8	0.01 (0.00)	0.89 (0.95)	0.330	0.000					

Note. The parameters of the asymptotic distributions are given within brackets.

(a) Information Function



(b) Bound for Error of Standard Normal Approximation



<u>Figure 1</u>. Information function and bound for error of standard normal approximation for subtest-residuals, in two cases of subtest: (1) of items with similar difficulties (dashed line), and (2) of items with distant difficulties (solid line).

Titles of recent Research Reports from the Division of Educational Measurement and Data Analysis. University of Twente, Enschede, The Netherlands.

RR-87-1 R. Engelen, *Semiparametric estimation in the Rasch model*

RR-87-2 W.J. van der Linden (Ed.), *IRT-based test construction*

RR-87-3 R. Engelen, P. Thommassen, & W. Vervaat, *Ignatov's theorem: A new and short proof*

RR-87-4 E. van der Burg, & J. de Leeuw, *Use of the multinomial jackknife and bootstrap in generalized nonlinear canonical correlation analysis*

RR-87-5 H. Kelderman, *Estimating a quasi-loglinear models for the Rasch table if the number of items is large*

RR-87-6 R. Engelen, *A review of different estimation procedures in the Rasch model*

RR-87-7 D.L. Knol & J.M.F. ten Berge, *Least-squares approximation of an improper by a proper correlation matrix using a semi-infinite convex program*

RR-87-8 E. van der Burg & J. de Leeuw, *Nonlinear canonical correlation analysis with k sets of variables*

RR-87-9 W.J. van der Linden, *Applications of decision theory to test-based decision making*

RR-87-10 W.J. van der Linden & E. Boekkooi-Timminga, *A maximin model for test design with practical constraints*

RR-88-1 E. van der Burg & J. de Leeuw, *Nonlinear redundancy analysis*

RR-88-2 W.J. van der Linden & J.J. Adema, *Algorithmic test design using classical item parameters*

RR-88-3 E. Boekkooi-Timminga, *A cluster-based method for test construction*

RR-88-4 J.J. Adema, *A note on solving large-scale zero-one programming problems*

RR-88-5 W.J. van der Linden, *Optimizing incomplete sample designs for item response model parameters*

RR-88-6 H.J. Vos, *The use of decision theory in the Minnesota Adaptive Instructional System*

RR-88-7 J.H.A.N. Rikers, *Towards an authoring system for item construction*

RR-88-8 R.J.H. Engelen, W.J. van der Linden, & S.J. Oosterloo, *Item information in the Rasch model*

RR-88-9 W.J. van der Linden & T.J.H.M. Eggen, *The Rasch model as a model for paired comparisons with an individual tie parameter*

RR-88-10 H. Kelderman & G. Macready, *Loglinear-latent-class models for detecting item bias*

RR-88-11 D.L. Knol & M.P.F. Berger, *Empirical comparison between factor analysis and item response models*

RR-88-12 E. van der Burg & G. Dijksterhuis, *Nonlinear canonical correlation analysis of multiway data*

RR-88-13 J. Kogut, *Asymptotic distribution of an IRT person fit index*

Research Reports can be obtained at costs from Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

department of

EDUCATION

A publication by
the Department of Education
of the University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands